DOCUMENT RESUME

ED 178 433 SO 012 097

AUTHOR Miller, Roy E.; Bositis, David A.
TITLE Exper-Sim vs. Setups: An Experiment-Based Comparative Evaluation Using Causal Modeling Techniques.
PUB DATE 77
NOTE 28p.; Paper prepared for delivery at Annual Meeting of the American Political Science Association (Washington, D.C., September 1-4, 1977)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Comparative Analysis; *Computer Assisted Instruction; Course Evaluation; Data Analysis; Educational Objectives; Graduate Students; Higher Education; Models; *Political Science; Research Methodology; Simulation; Statistical Analysis; Student Attitudes; Student Motivation; *Teaching Methods

ABSTRACT

Two approaches for teaching empirical research methods to graduate students in political science courses are evaluated. Specific objectives of the evaluation, in addition to comparing relative teaching effectiveness of the two approaches, were to explore causal dynamics between the teaching environment, user satisfaction, and learning performance. The first approach, EXPER-SIM (Experiment Simulation), is a computer program which transmits simulation models into fortran subroutines and generates data according to a selected model. The second approach, SETUPS, (Supplementary Empirical Teaching Units in Political Science), includes instructional packages covering various aspects of political science. A SETUP typically contains an introduction to theory and literature, data analysis exercises, codebook and data sets, and directions for use in conjunction with statistical software packages. Evaluation of the approaches took place during a one-semester course in which students were randomly assigned to either EXPER-SIM or SETUPS groups. Students in both groups investigated the same analytic problems. Findings from statistical analysis of pre-and posttest scores and from general course ratings by students indicated that the EXPER-SIM students showed greater gains in analytical reasoning ability, positive perceptions of computer utility, ease of computer use, and substantive learning performance. The conclusion is that EXPER-SIM is more effective than SETUPS as a strategy to teach methods of empirical research. (DB)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

ED178433

# EXPER-SIM VS. SETUPS: AN EXPERIMENT-BASED COMPARATIVE EVALUATION USING CAUSAL MODELING TECHNIQUES

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Roy E. Miller
David A. Bositis
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

By

Roy E. Miller

and

David A. Bositis

Department of Political Science

Southern Illinois University at Carbondale
Carbondale, Illinois

SO 012 097

Prepared for delivery at the 1977 Annual Meeting of the American Political Science Association, The Washington Hilton Hotel, Washington, D.C., September 1-4, 1977. Copyright by The American Political Science Association, 1977.

# EXPER-SIM VS. SETUPS: AN EXPERIMENT-BASED COMPARATIVE EVALUATION USING CAUSAL MODELING TECHNIQUES*

## Introduction

This paper presents some findings from an experimental evaluation of the comparative teaching effectiveness of two relatively new active teaching technologies, EXPER-SIM (Experiment Simulation) and SETUPS (Supplementary Empirical Teaching Units in Political Science), in a required graduate-level "Methods of Empirical Research" class in political science.[1]

EXPER-SIM (Main, 1974) is a sophisticated fortran simulation supervisor program developed at the University of Michigan which permits the programming of simulation models as fortran subroutines (Stout, 1974).[2] Once operational, the system then generates data according to the selected model within the context of user-designed simulated experiments. Up to this time, most of the models developed for EXPER-SIM have been simulations of psychological processes (Main, 1971). However, EXPER-SIM is capable of simulating experiments with substantive models from any discipline, and as the supervisor system has been implemented at other universities, models in some other disciplines have been developed.[3]

SETUPS, developed through the joint efforts of the American Political Science Association and the Inter-University Consortium for Political and Social Research, includes a number of data-based instructional packages covering various

---

*This study was made possible by a grant from the IMPACT Program of the Exxon Education Foundation. In addition, we should like to acknowledge our indebtedness to Fred Ribich for his technical assistance in the implementation of EXPER-SIM at SIU, and to the students whose unknowing cooperation was instrumental to the successful completion of the study -- many thanks to Asukwo, Allan, Barbara, Corrin, Hiroko, Jane, Jay, Mariana, Myung, Rodney, and Sylvia.

areas in political science. Each individual SETUPS typically contains an introduction to the theory and literature prevalent in the given substantive area, several data-analytic exercises, a codebook and a data set, and is intended for use with one of several statistical software packages (such as SPSS or OSIRIS) in the batch processing mode.

For the experiment being described herein, the particular SETUPS used was Voting Behavior: The 1972 Election (Bowen, Broh and Prysby, 1975), and the students using SETUPS did so in conjunction with SPSS operating in the batch processing mode. The students using EXPER-SIM, on the other hand, explored the same substantive area of "voting behavior in the 1972 election" by means of a simulation model (VOTE72) whose parameters had been estimated from analysis of the aforementioned SETUPS data set, and which was accessed in the interactive mode from remote terminals.[4]

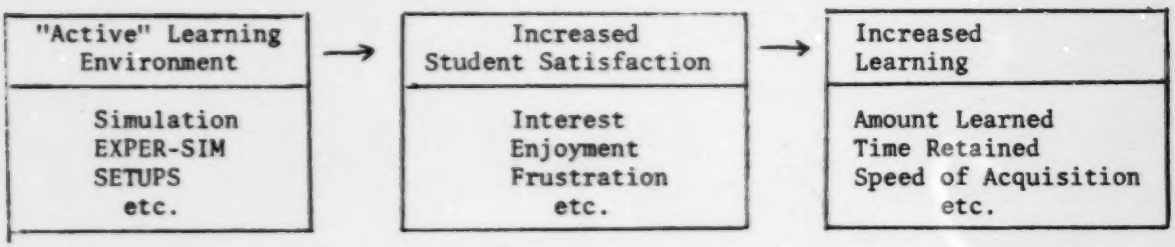
While the use of EXPER-SIM and SETUPS/SPSS are both "active" teaching environments when compared to the more "passive" traditional lecture format, EXPER-SIM seemingly is the more "active" of the two. Since EXPER-SIM operates in the interactive mode,[5] it provides an immediate response to the commands entered by the user, who is then stimulated to respond fairly rapidly to the data or questions generated by the system. On the other hand, the use of SETUPS with a software package such as SPSS necessarily entails the normal thru-put time lags endemic to batch processing.

It was this "activeness" difference which we believed would be responsible for any differential group performance resulting from the use of EXPER-SIM in the interactive mode versus SETUPS/SPSS in the batch processing mode.[6] In particular, we believed that the more "active" nature of EXPER-SIM would serve as a stimulus to heightened student interest, and hopefully in turn, increased student performance. A number of studies have found the former to be true. For example, Rywick (1975) found a significantly greater degree of student interest using interactive

computer simulations when compared to person simulations. Likewise, Johnson (1974) found heightened student interest with computer simulation in the interactive mode.[7] On the other hand, while the majority of research has been disappointing in its inability to demonstrate a significant relationship between learning performance and teaching technology (e.g., Kissler, 1974; Rywick, 1975), there is a limited amount of evidence which does suggest that the use of a more active technology results in increased learning performace (Broh, 1976; Nussloch, Kaplan and Main, 1974).

Most important, however, is the fact that at the present moment in time there seems to be no clear and convincing evidence that a causal dynamic exists which flows from teaching technology to student satisfaction to learning performance, as diagrammed in the causal model of Figure 1.

Figure 1. A Causal Model of A Basic Assumption Made By Many CAI Advocates




In simple terms, the model states that the more "active" learning environment will produce "happier" students who will therefore learn "better."

The learning environment can be made a more "active" one in many ways, for example, by the use of man-man, man-computer, or computer simulation, role playing, gaming, doing data-analytic exercises with the aid of a computer, and so forth. Likewise, there may be many different indicators of "student satisfaction," such as interest, enjoyment, frustration, perceived difficulty, perceived usefulness, preference, and so on. And, of course, there are many different indicators of what might constitute a "better" learning performance, such as the amount of material learned in a given amount of time, the time

required to learn a given amount of material, the length of time that information is retained, the type of material that is learned (some things are more important than others), and so forth. The relative failure of past "teaching evaluation" research to clearly substantiate the causal dynamic portrayed in Figure 1 seemingly could be due to one or more of several reasons.

First, much of the evaluation research concerning the effectiveness of "active" teaching environments has focused exclusively on the relationship between the learning environment and indicators of student satisfaction (e.g., Parker and Ross, 1977). Secondly, much of the evaluation research has utilized what might be thought of as relatively "soft" measures of learning performance, in particular attitudinal measures of what students "thought" they had learned (e.g., Fiedler, 1976). Thirdly, seemingly most of the evaluation literature in this area has either relied exclusively upon a bivariate logic or has utilized analytic techniques that are less than optimal for investigating the structure and process of a multivariate causal system. Or, lastly, perhaps it is the case that the assumed basic causal dynamic is simply false.

Given the causal assumption portrayed in Figure 1, but realizing that there is no clear and convincing empirical evidence supporting that assumption, the objectives of this paper can be stated as follows: (1) to comparatively evaluate the relative teaching effectiveness of the use of EXPER-SIM versus SETUPS/SPSS, and (2) to make such an evaluation using multivariate logic and standard causal modeling techniques (Blalock, 1961 and 1971) so as to facilitate the elucidation of the causal dynamic involved between teaching environments and learning performance.

## Teaching Goals And Indicators

Before one can evaluate the comparative teaching effectiveness of any two

teaching strategies or technologies, one must first address the question of "effectiveness in terms of what?" In other words, in any goal-oriented evaluation (in contrast to goal-free evaluation) the basic question may be phrased as follows: "Is program A doing better than program B in reaching their common goals?" (Weiss, 1972:24) In order to bring evidence to bear on this question, one must first clearly specify the program goals that one hopes will be achieved.

Typically, teachers have multiple goals in the classroom situation, and ordinarily at least some of the goals are not independent of one another; thus, they may not necessarily be complimentary. For example, in a class in "Methods of Empirical Research" the teacher may wish to maximize the degree of competence that students acquire with respect to being able to productively use the computer in their research activities, while at the same time wish to minimize the degree of frustration or difficulty that students have in their use of the computer. The accomplishment of the former goal would seem to imply extensive use of a computer by the students, whereas accomplishment of the latter goal might imply only minimal use of a computer. The class which provided the data for the present work was no exception. As teachers, we indeed had multiple goals, some of which may not have been complimentary. Our goals, and their relative priorities, may or may not be isomorphic with the goals and priorities of other courses or instructors. Regardless, the identification of those goals and priorities is necessary before we can proceed with a comparative evaluation of the two teaching technologies -- EXPER-SIM and SETUPS/SPSS.

Like most teachers, our goals for this course included goals in both the user satisfaction and learning performance domains. Paramount, however, were the goals in the learning performance domain. In that domain, we developed "hard" measures of four substantive components of the course, those being "analytical reasoning ability," "computer knowledge," "statistical knowledge," and "general 'methods' knowledge." For the purposes of this paper, only the

"analytical reasoning" component will be discussed.[8] We do this because of time and space constraints, and because it is this component that we clearly considered to be the most important of the four.

The set of analytical reasoning questions focused on the students' ability to respond to problems which generally required a logical and stepwise solution. Some of these questions required the student to proceed from a set of given assumptions and information to solution of a problem. Others required students to interpret concepts, which they had been exposed to in other contexts, but in a slightly different fashion. Still others required quantitative analysis, sometimes based upon either tabular or graphic presentations of data, within a social science context. The following question is illustrative of the items used to measure "analytical reasoning ability."

> We have a causal law. The voter, who can vote either Democratic (D) or Republican (R) is the subject of this law. As the past, let us consider the last three elections. The law states that if, in the last three elections, the voter voted Democratic exactly twice, then the voter votes the same in the next election as in the previous election. Otherwise, the voter will vote for the other party.
>
> Suppose we have a voter who voted Democratic, Democratic, and Republican in the last three elections. Let us symbolize this by DDR. How will the voter vote in the next four elections?
>
> a. DRRD
> b. RDDR
> c. RDRD
> d. RRRD

Turning now to the "user satisfaction" domain, our goals in this domain were again multiple in number, but our measured indicators of these goals were necessarily of the "softer" variety. By definition, they were goals involving student perceptions and/or opinions. In addition, we were somewhat less clear and certain as to the relative priorities among our user satisfaction goals.

Because the objectives of this paper are in part focused on the nature and configuration of the causal process (if any) as it connects the teaching environment with learning performance through the intervening attitudinal domain of user satisfaction, all five of our user satisfaction measures will be included in this analysis. All of our user satisfaction measures were of the Likert scale type and consisted of attitudinal questions which were neutral or non-specific in their wording (that is, with regard to either EXPER-SIM or SETUPS/SPSS) and which utilized a five-point response scale ranging from strongly agree to strongly disagree.

The two user satisfaction goals which we considered most important were (1) that students should have minimal difficulties in actually using the computer, and (2) that the course should increase the students' appreciation of the potential usefulness of the computer in data analysis and research. These two goals are actually intermediate goals, the attainment of which we would hope would insure that in their future careers the students would actually use the computer whenever such usage would be advantageous to them. And, of course, students are much more likely to actually use the computer if they find it relatively easy to use and if they perceive that the computer is a useful research tool. Illustrative examples of the attitudinal items used to measure "Perceived Difficulty of Computer Use" and "Perceived Computer Utility" are as follows.

(1) Generally I have not had much trouble getting the computer to do what I asked of it.

(2) I now see the computer as a much more useful "research tool" than I did when I began this course.

Another goal of perhaps lesser importance than those just discussed was that of optimally matching the difficulty level of the course with the level of competence and understanding of the students. The achievement of this

goal was desirable to insure that students did not spend all of their time "cramming," or rushing to finish this work or that assignment, or were constantly confused; but rather would have sufficient time and conceptual clarity to reflect on the meaningful subtleties and intricacies which are the very lifeblood of the empirical analysis of social phenomena. "Perceived Course Difficulty" was measured using the standard set of items taken from the "Instructional Improvement Questionnaire" (IIQ) used at Southern Illinois University, of which the following item is an illustrative example: "The course material was too difficult."

A fourth goal, of probably about the same importance as that just discussed, was the goal of increasing students' interest in and enthusiasm for the study of the substantive course material itself. This was desirable for two reasons: (1) because of the assumption discussed earlier that increased student interest would result in increased learning performance, and (2) regardless of the validity of that assumption, because of the instructors' strong normative belief in the importance of the subject matter vis-à-vis the adequate preparation of any graduate student in political science. The following is illustrative of the items used to measure student "Interest and Enthusiasm:" "I'm just not very interested in learning about empirical data analysis."

Our final goal, and most certainly the least important among our user satisfaction goals, was that of achieving a fairly high overall student evaluation of the course. Again, this was desired for basically two reasons. First, we assume that student course ratings reflect the overall satisfaction that the direct recipients (students) of our teaching services have with the quality of those services; and, like most other teachers, we like to be told we are doing a "good job." Secondly, it is a contemporary fact of life at SIU and at many other universities that student course evaluations play a role

in administrative decisions regarding faculty tenure, promotion, and salary questions; therefore, high course ratings may be desirable to instructors from a simple pragmatic perspective. The "General Course Rating" of students was again measured using the standard set of items taken from the IIQ, of which the following is an illustrative example: "This course was one of the best I have taken."

In sum, our goals for this course were multiple in number and of differential priorities. Our most important goals were learning performance goals, with student satisfaction goals being of lesser import. "Hard" learning performance measures were developed, but only that concerning "Analytical Reasoning Ability" will be investigated in this paper. On the other hand, all five of our student satisfaction measures (Perceived Difficulty of Computer Use, Perceived Computer Utility, Perceived Course Difficulty, Interest and Enthusiasm, and General Course Rating) will be included in our analysis.

## Research Design And Measurement Procedures

The compartive evaluation of the effectiveness of the use of EXPER-SIM versus SETUPS/SPSS is based upon data gathered during a semester-long course using a counterbalanced experimental design with random assignment of both subjects and treatments to groups (Campbell and Stanley, 1966). During each half of the semester, the experimental "treatment" lasted six weeks and consisted of the students doing weekly data analytic lab assignments using either EXPER-SIM or SETUPS/SPSS. Effectively, each week's lab assignment was worded identically for the two groups, so on a week to week basis the students in both treatment groups were being asked to investigate the same analytic problems. The only thing that differed was the fact that one group of students was investigating those problems in the interactive mode via the use of EXPER-SIM

and the VOTE72 model, whereas the other group of students was exploring those same problems in the batch mode using the SETUPS documentation and SPSS. The design for the entire semester is portrayed in Figure 2.

Figure 2. Research Design Used For The Comparative Evaluation Of EXPER-SIM Versus SETUPS/SPSS

| | Time1 Beginning Of Semester | Experimental Treatment | Time2 Mid-Semester | Experimental Treatment | Time3 End Of Semester |
|---|---|---|---|---|---|
| [R]* Group1 (N=5) | Pretest | Assignments Using EXPER-SIM | Posttest1 | Assignments Using SETUPS/SPSS | Posttest2 |
| Group2 (N=5) | Pretest | Assignments Using SETUPS/SPSS | Posttest1 | Assignments Using EXPER-SIM | Posttest2 |

[R]* indicates random assignemt of both subjects and treatments to groups.

It can be quickly seen from the design that in effect we sequentially ran two complete experiments in the same course during the same semester with the same students as subjects, with both experiments utilizing a Pretest-Posttest Two-Group design. Given that findings from the data gathered during the first half of the semester have been reported elsewhere (Miller and Bositis, 1977), we shall limit our attention in this paper to an analysis of the data generated from the experiment conducted during the second half of the semester. In so doing, we will be considering the measures taken at Time2 (Mid-semester) as pretest measures, and the measures taken at Time3 (end of semester) as posttest measures.

Some brief comments are in order with regard to the validity of the research design which we have used. As Campbell and Stanley (1966) have pointed out, the Pretest-Posttest Two-Group design is very strong in its control of standard threats to the internal validity of one's findings, such as history,

maturation, testing, selection, and so forth. With respect to external validity, however, the design is somewhat deficient, in particular its inability to control for the potential interaction of pretesting and the treatment variable. We do not consider this potential interaction threat a very serious one, however, given that what served as a pretest (for the purposes of this paper) was the mid-term examination, and by and large students are probably quite used to having both mid-term and final examinations. It should be noted, though, that even more important than this potential interaction threat is the fact that the external generalizability of our findings is necessarily constrained by many other factors. It is a fact that this experiment was done in only one class, at only one university, in only one section of the country, with a non-randomly selected sample of students, etc., etc., etc. Like most experimental research, the sample, ecological, and variable representativeness of our study is indeed open to question, and, again like most experimental research, we can only hope that the external generalizability of our findings is strengthened by the process of replication in other settings and by other researchers.

It now needs to be made explicitly clear that the causal modeling analysis which is to follow actually makes use of Time3-Time2 "difference" scores -- not raw scores from either time period alone. We have used difference scores for basically two reasons. First, with regard to the experiment conducted during the second half of the semester we absolutely knew that our two groups were non-equivalent at the start of that experiment (mid-semester) in at least one respect: Group1 had just used EXPER-SIM for six weeks, whereas Group2 had just used SETUPS/SPSS for six weeks. By using difference scores in our analysis, we effectively control for this known pre-experimental non-equivalence between our groups. Secondly, a basic assumption of causal modeling is that no variables outside the system under investigation act upon any of the endogenous variables in the system in a manner which is systematically related

to any of the exogenous variables in the system (Blalock, 1961). One of the extremely pleasant aspects of doing causal modeling with data generated from the use of randomized experimental designs is the fact that the random assignment of subjects to experimental groups is normally sufficient to allow the researcher to make the above assumption -- at least with respect to the all important experimental treatment variable. When one is faced with a relatively small number of subjects however, as was the case in this experiment, the use of difference scores augments the randomization procedures, and in so doing buttresses the above assumption.

Before proceeding, mention should be made of an additional pre-analysis adjustment which was performed on our measure of "Analytical Reasoning Ability," namely, the weighting of the items that were used in that measure. It will be recalled that unlike our attitudinal user sacisfaction measures, our test of "Analytical Reasoning Ability" was made up of objective questions to which there were right and wrong answers. By design, some of the items included in this test were more difficult than others. In order to insure that scores on this learning performance measure were sensitive to this variability in item difficulty, each item was weighted by the reciprocal of the number of students who answered that item correctly. Accordingly, if only one student in the class answered a particularly difficult item correctly, the weighted contribution of that item to the total test score would be one (1); whereas if nine of the ten students answered an easy item correctly, the weighted contribution of that item to a student's total test score would be only one-ninth (1/9).

In summary of this section, we have indicated that the research design used in this study was a two-group counterbalanced experimental design, that we shall limit our analysis in this paper to data that was generated during the second half of the semester, and that while the research design used

has some deficiencies vis-ā-vis its external validity, it is quite strong with respect to its internal validity. In addition, we have made it clear that the forthcoming causal modeling analysis is based upon the use of Time3-Time2 difference scores, and that our measure of "Analytical Reasoning Ability" is a weighted measure so as to take into consideration the degree of item difficulty. We shall now turn our attention to the causal modeling analysis itself.

EXPER-SIM Versus SETUPS/SPSS: The Development Of A Causal Model

At the outset, it should be stipulated that our modeling efforts were limited to the exploration of "recursive" models. In any recursive causal system of seven variables there will be 21 pairs of relationships, and, since any given causal arrow in such a system may be either present or absent, there will be $2^{21}$, or 2,097,152, total possible causal configurations. Quite obviously, we did not, and could not, examine each of the possible causal models to see how well it fit our data -- although we did examine a great many. Also obvious is the fact that it is quite probable that several of the possible models would show an equally acceptable empirical fit with our data, although not necessarily an equally acceptable conceptual or theoretical fit. The point is that the reader should explicitly understand that the model presented herein was guided in its development by our theoretical expectations regarding how the system in question ought to operate, as well as by our previous analysis of the data gathered during the first half of the semester (Miller and Bositis, 1977). Accordingly, the model presented in Figure 2 is not necessarily the "best" fitting model of all possible models of the system; but, rather, is a "good" fitting model in terms of both empirical expectations and theoretical

understandings -- and, indeed, is the best fitting of all the numerous models that we examined.

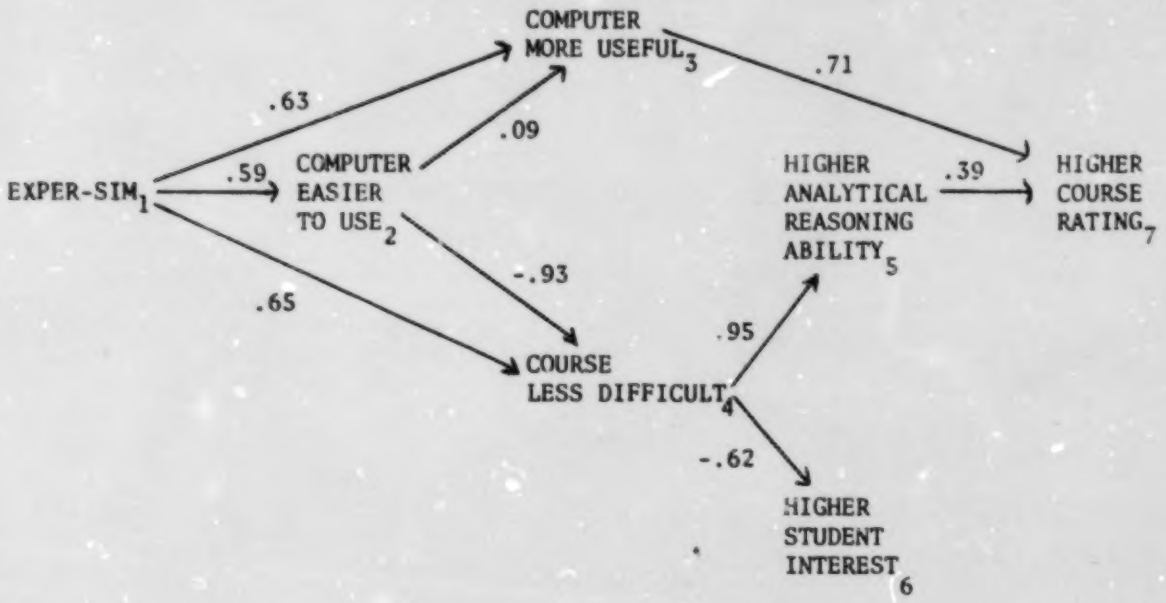
**** Figure Here ****

Before proceeding to a substantive discussion of the model presented in Figure 3, a few clarifying comments about that figure might be useful. To begin with, in an attempt to hopefully facilitate the reader's quick and clear understanding of the directional implications of the causal linkages, all of the "variables" in the model have been stated with directional qualifiers; thus the variable "Perceived Computer Utility" appears in the model as "Computer More Useful," and the variable "Learning Environment" appears as "EXPER-SIM." Accordingly, the direct positive path between these two variables indicates that the students who used EXPER-SIM perceived the computer to be more useful. Secondly, the coefficients appearing in the model are standardized path coefficients.[9] And, lastly, accompanying the model is a minimal set of "goodness of fit" tests, i.e., a series of empirical expectations involving low-order partial correlations that the model implies.[10] It is obvious that even if one were to use the more liberal .10 alpha level, in no instance does the data from our experiment significantly differ from the model expectations.[11] In other words, the model has a good empirical fit with the data.

At this point, the reader should be made explicitly aware that the following substantive discussion of the model and its development reflects what Kaplan (1964) calls a "reconstructed logic." The written narrative may make it appear as though the model development was a neat, step-by-step, one variable at a time sort of process. In reality, of course, the actual developmental process was not nearly so tidy, and at almost all points model-fitting decisions were based upon multivariate considerations.

Given that we utilized an experimental research design, which means that

Figure 3. A Model Of The Causal Dynamics Linking The Teaching Environment With Learning Performance Via The Intervening Attitudinal Domain Of User Satisfaction




A Minimal Set of "Goodness Of Fit" Tests Using Model Expectations Involving Low-Order Vanishing Partials

| Model Expectation | Empirical Result | Significance Level (One-Tailed Test) |
|---|---|---|
| $r_{15.4} = 0$ | -.35 | .18 |
| $r_{16.4} = 0$ | -.24 | .2[illegible] |
| $r_{25.4} = 0$ | -.16 | .34 |
| $r_{26.4} = 0$ | -.22 | .28 |
| $r_{35.4} = 0$ | -.03 | .47 |
| $r_{36.4} = 0$ | .00 | .50 |
| $r_{56.4} = 0$ | .19 | .31 |
| $r_{67.4} = 0$ | .19 | .31 |
| $r_{17.35} = 0$ | -.49 | .11 |
| $r_{27.35} = 0$ | -.15 | .37 |
| $r_{34.12} = 0$ | -.31 | .23 |
| $r_{47.35} = 0$ | .08 | .43 |

we randomly assigned subjects to treatment groups long before the measurements on our other variables were taken, we were assured that the single exogenous variable in the system was the nature of the learning environment, i.e., EXPER-SIM versus SETUPS/SPSS. On the basis of analysis of the mid-semester data, we had clear evidence that once one controlled for pre-experimental attitudes and competences the variable that was the most strongly related to the treatment variable was student perceptions of the "Difficulty of Computer Use." The data from the second half of the semester supported this finding, and thus the model begins with a direct positive path indicating that the students using EXPER-SIM found the computer easier to use. This link is perhaps the cornerstone of the model, and is the link of which we are the most certain.

The next variable fitted to the model was "Perceived Computer Utility." The model indicates that there is both a direct positive path of influence from the use of EXPER-SIM to the students' perception of the utility of the computer, as well as an indirect positive path through the influence of the students' perceptions of how easy the computer is to use. This indirect path seems conceptually quite reasonable since it makes good sense that if someone sees the computer as relatively easy to use they would likewise see the computer as being potentially more useful.

The next variable brought into the model was "Perceived Course Difficulty." Looking first at the direct path from EXPER-SIM to "Course Difficulty," we find that the use of EXPER-SIM resulted directly in perceiving the course as being less difficult. At the same time, however, the use of EXPER-SIM produces a contrary effect on "Course Difficulty" via the indirect path through "Difficulty of Computer Use." Initially, this indirect path seemed rather curious, but, upon further thought, it made sense that when students find the computer easier to use they will spend less time on the more mechanical computer processing

portion of the course and will increase their concentration and time spent on the essential conceptual difficulties posed in data analysis (which comparatively speaking is the more difficult of the two). When one considers the relative magnitude of these two paths, the net effect is that students using EXPER-SIM perceived the course as being slightly less difficult.

The next variable entered in the model was "Interest and Enthusiasm." The fitting of this variable to the model was somewhat problematic. To begin with, proceeding from the basic assumption diagrammed in Figure 1, our initial expectation was that there should be a direct link between the nature of the learning environment and the degree of student "Interest and Enthusiasm." However, the data simply did not support such a direct link. In fact, we could find no acceptable way to add "Interest and Enthusiasm" to the model except via a link with "Course Difficulty," and the causal direction of that link was somewhat indeterminate. From the perspective of the empirical "goodness of fit" requirements, "Interest and Enthusiasm" could be added to the model equally well as either a cause of or an effect of "Course Difficulty." Accordingly, given the negative sign of the path between these two variables, we have chosen to enter "Interest and Enthusiasm" into the model as an effect of "Perceived Course Difficulty," because that positioning seemed to make more conceptual sense to us. That is to say, it seems more reasonable to conclude that as students see a course as becoming increasingly easier they might tend to lose some interest in the course, than it does to suggest that students losing interest in a course causes them to see it as being easier.

We next entered the last of our user satisfaction measures, namely, "General Course Rating." For this variable, there seemed to be no acceptable fit to the model except as an effect of "Perceived Computer Usefulness" with students who saw the computer as a more useful research tool tending to give

the course a higher rating. This particular link had been anticipated, since it seemed quite reasonable to expect that students would rate a course more favorable if they thought they were being taught how to use something that they perceived as being potentially useful. However, we might point out that we certainly did not expect this to be the only direct causal path between "Course Rating" and any of the user satisfaction measures; in particular, the lack of any direct link between student "Interest and Enthusiasm" and "General Course Rating" was quite surprising.

Lastly, we entered the variable of "Analytical Reasoning Ability," and after much probing concluded that the only place it would acceptably fit with the model was as an intervening variable between "Course Difficulty" and "Course Rating." It was at this point that our basic assumption about the chronological placement of the teaching environment, user satisfaction, and learning performance measures had to be re-examined, since contrary to our expectations our measure of learning performance was not the ultimate dependent variable in the model. As we thought about this situation, our attention became focused on our "General Course Rating" measure. All of our other user satisfaction measures are of fairly specific attitudinal states, whereas "General Course Rating" seems to be an attitudinal measure with a much less specific referent. In fact, it can be argued that a student's "Course Rating" is really a composite measure reflecting many things, including perhaps such things as the student's personal feelings about the instructor, the student's expected grade in the course, the size of the class, the time at which the class meets, whether or not the course is a required course, and, most importantly, whether or not the student actually learns anything in the course. Having now conceptualized our "Course Rating" measure

as being slightly different from our other user satisfaction measures, and having now realized that how much a student learns from a course is (hopefully) a partial determinant of how favorably a student rates that course, the fact that our learning performance measure comes before our "Course Rating" measure in the model now seems quite reasonable.

Looking now at the direction of the paths leading to and emanating from our measure of "Analytical Reasoning Ability," one sees that students who perceived the course as being less difficult demonstrated a higher increase in "Analytical Reasoning Ability," and that a higher increase in "Analytical Reasoning Ability" lead to higher course ratings. The path from "Course Difficulty" to "Analytical Reasoning Ability" at first blush may seem somewhat mystifying in terms of its sign. Hearkening back to an earlier discussion, however, it will be recalled that in explaining the sign of the path from "Difficulty of Computer Use" to "Perceived Course Difficulty" it was reasoned that when students found the computer easier to use they would spend less time on the more mechanical computer processing portion of the course and would increase their concentration and time spent on the conceptual difficulties posed in data analysis. And, therefore, by increasing their concentration and time spent on the more analytical component of the course, it would seem quite reasonable that students would show improvement in their analytical reasoning powers.

The direct positive path from "Analytical Reasoning Ability" to "Course Rating" indicates that as students improved on the analytical reasoning measure they tended to rate the course higher. It would seem that the sign and magnitude of this path would be a function of at least two factors. First, again hearkening back to an earlier discussion, it seems to us that generally speaking course ratings can be expected to reflect to some degree the magnitude of

learning that occurs in a course, i.e., students will give higher ratings to courses in which they actually learn something. Secondly, however, the particular type of learning performance being measured here (analytical reasoning ability) is perhaps quite course specific, and it does not seem reasonable to us to always expect an improvement in learning performance of any type to necessarily lead to higher course ratings. Then too, if one only considers this particular type of learning performance, it would seem reasonable that as students increase their analytical reasoning abilities they would utilize these improved critical faculties in evaluating their learning environment when course-rating time came around. Thusly, we can easily envision the situation where students with increased analytical abilities might critically evaluate their learning environment and find it seriously lacking for any number of reasons, and hence give the course lower (rather than higher) ratings. In this experiment, increased analytical reasoning abilities did lead to higher course ratings, but this is quite possibly a situational specific finding.

To sum up, the overall model presents the following picture of the relative advantages of the use of EXPER-SIM versus SETUPS/SPSS: (1) The EXPER-SIM students definitely found the computer easier to use which in turn contributed to their seeing the computer as being a more useful research tool. (2) The EXPER-SIM students perceived the course as being somewhat easier which contributed to both declining student interest and improved learning performance (analytical reasoning ability). (3) Seemingly because of their perceptions of greater computer utility and their improved analytical reasoning abilities, the EXPER-SIM students rated the course more favorably than did the SETUPS/SPSS students. Accordingly, with respect to five out of six teaching goals that we had specified for this course, including our three most important goals, EXPER-SIM seems to

be the more effective teaching technology to use. The use of SETUPS/SPSS would seem to the more desirable strategy vis-à-vis the attainment of only one of our six goals, that being to increase student interest and enthusiasm -- a goal which was low on our list of priorities.

## Conclusion

We began this paper with two objectives in mind: (1) to comparatively evaluate the relative teaching effectiveness of the use of EXPER-SIM versus SETUPS/SPSS, and (2) to do so using causal modeling techniques so as to explore the causal dynamics involved between the nature of the teaching environment, user satisfaction, and learning performance. With respect to the first objective, given our teaching goals and their priorities, the evidence we have presented clearly suggests that the use of EXPER-SIM is the preferred teaching strategy. Its use seems to facilitate the attainment of five out of our six specified teaching goals, including our three most important ones -- (1) maximizing substantive learning performance, (2) minimizing the difficulties that students have in actually using the computer, and (3) maximizing the students' appreciation of the potential usefulness of the computer as a research tool.

With respect to the second objective, we have developed and presented a model which (1) successfully incorporates all seven of the variables with which we were concerned, (2) shows a reasonably good fit with our empirical data, and (3) generally makes sense in terms of both its theoretical and conceptual implications as well as the behavioral nuances we observed throughout the semester-long process of teaching and collecting evidence. If one assumes that the model is somewhat isomorphic with reality, then the model fruitfully and synoptically demonstrates that in the specific context

of this course the nature of the teaching environment did indeed affect learning performance via the intervening attitudinal domain of "user satisfaction." In particular, the model is generally consonant with and lends empirical support to the basic assumption with which this paper was begun, namely, that the more "active" learning environment increases student satisfaction which results in increased learning.

However, our model, and therefore our data, does indicate that one of the most often used measures of student satisfaction, namely, course ratings, is not a part of the intervening attitudinal domain which facilitates increased learning. In other words, course ratings are partially dependent upon whether or not students learn anything in a course, not vice versa. What this implies is that courses with high course ratings are not necessarily courses in which students are learning a lot, and, therefore, that information about the former (course rating) should not be used as if it were necessarily information about the latter (amount of learning that has occurred).

In addition, we should here highlight the fact that our model and data do not support the often expressed idea that increasing student interest in a course will lead to increased learning performance, but, rather the model indicates that in this course learning performance was in no way dependent upon student interest. With respect to this finding, two comments would seem to be in order. First, this experiment was done in a course that students were required to take, and it is quite possible that the causal relationship between student "Interest and Enthusiasm" and learning performance is different for required courses than it is for elective courses. Second, it seems to us that there is a conceptual difference between "interest" in a subject matter and "motivation" to learn about a subject matter so as to understand it

in all of its richness and complexity and be able to apply it in one's professional life. The model points out that an increase in student "interest" does not necessarily lead to increased learning, but the model says nothing about the relationship between "motivation" and learning.

Quite obviously, other teachers and other courses may have different goals from what we have been concerned with and/or different priorities. The findings presented herein are based upon an experiment done in a single class within a single environmental context. Within those constraints, our findings generally favor the use of EXPER-SIM over SETUPS/SPSS. Hopefully, other teachers in other settings with different goals and/or priorities will experimentally incorporate the use of EXPER-SIM into their teaching, systematically evaluate its effectiveness, and add their findings to a cumulative body of information concerning EXPER-SIM. Only then will we be able to draw a clearer and more complete picture of the strengths and weaknesses of the use of EXPER-SIM as a teaching strategy.

## NOTES

1. The particular class involved was POLS 501-B, taught during the Spring Semester, 1977, at Southern Illinois University at Carbondale.

2. There are other experimental simulation systems available, most notably, DATACALL (R. Johnson, 1971), LABSIM (Kissler, 1974), and LESS (Hedberg, 1976).

3. For example, the Political Science Laboratory at Ohio State University has developed a "social choice" model.

4. The VOTE72 simulation model was designed, programmed and implemented at SIU by Dr. Roy E. Miller in December, 1976.

5. It should be noted that EXPER-SIM can also be used in a batch-processing environment at those computer installations that do not have time-sharing capabilities.

6. For an extensive discussion of all of the expectations which we had at the start of the experiment, as well as their supporting rationale, see Miller and Bositis (1977).

7. A note of caution from G. Kissler (1974) is that the initial superiority of on-line simulation may be due to the novelty of using remote terminals.

8. For a discussion of the other goals in the learning performance domain, as well as some initial findings pertinent thereto, see Miller and Bositis (1977).

9. Standardized path coefficients are reported here because of their ease of interpretation, their lack of dependence upon the variance of one's measurement scales, and their utility for comparing the relative magnitudes of different paths within a given model for a given sample/population. For a discussion of path analysis generally, see Part II in Blalock (1971).

10. It should be noted, of course, that the model actually implies numerous other statistical expectations including partials that are expected to be zero, partials that are expected to be non-zero, and partials that are expected to be less than zero-order coefficients.

11. Out of a total of 112 "goodness of fit" tests, in only 5 instances did our data not match model expectations.

## REFERENCES


Blalock, H. M., Jr. (1961) Causal Inferences in Nonexperimental Research. Chapel Hill, North Carolina: University of North Carolina Press.

_____ (1971) Causal Models in the Social Sciences, ed. Chicago: Aldine.

Bowen, B. D., C. A. Broh, and C. L. Prysby (1975) Voting Behavior: The 1972 Election. Washington: American Political Science Association.

Broh, C. A. (1976) "Using SETUPS: American Politics -- A Field Experiment." DEA News 9. (A publication of the Division of Educational Affairs of the American Political Science Association)

Campbell, D. T. and J. C. Stanley (1966) Experimental And Quasi-Experimental Designs For Research. Chicago: Rand McNally.

Cromer, A. (1974) "Teaching Research Theory and Application through On-Line Simulation." Behavior Research Methods and Instrumentation 6.

Fiedler, M. (1976) "SETUPS: American Politics Field Tests -- First Reactions: Positive, Enthusiastic." DEA News 9. (A publication of the Division of Educational Affairs of the American Political Science Association)

Halstead-Nussloch, R. H. and D. B. Main (1975) "EXPER-SIM Pedagogy: Inference of Models thorugh Research Design Instruction." Paper presented at the Annual Meeting of the American Psychological Association, Chicago, Illinois.

Hedberg, J. G. (1976) "The Technologies of EXPER-SIM." Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, California.

Johnson, R. (1971) "DATACALL: A Computer Based Simulation Game for Teaching Strategy in Sientific Research." Paper presented at the Second Conference on Computers in the Undergraduate Curricula, Dartmouth College, Hanover, New Hampshire.

_____ (1974) "Instructional Simulation: The Interface with the Student." Behavior Research Methods and Instrumentation 6.

Kaplan, A. (1964) The Conduct of Inquiry. San Francisco: Chandler.

Kissler, G. R. (1974) "Evaluation of Computer-Based Laboratory Simulation Models to Teach Scientific Research Strategies." Behavior Research Methods and Instrumentation 6.

Lehman, R. S., B. J. Starr, and K. C. Young (1975) "Computer Aids in Teaching Statistics and Methodology." Behavior Research Methods and Instrumentation 7.

Main, D. B. (1971) "Computer Simulations in the Elementary Psychology Laboratory." Paper presented at the Second Conference on Computers in the Undergraduate Curricula, Dartmouth College, Hanover, New Hampshire.

_____ (1972) "Toward A Future-Oriented Curriculum." American Psychologist 27.

_____ (1974) EXPER-SIM on Michigan Experimental Simulation Supervisor, ed. Columbus, Ohio: Xerox.

Miller, R. E. and D. A. Bositis (1977) "EXPER-SIM Versus SETUPS: Some Initial Findings." Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago, Illinois.

Nussloch, R. H., R. J. Kaplan, and D. B. Main (1974) "Casual Critiques or Active Arguments? -- A Comparison of the Use of Computer Data Generators and Journal Article Criticism in the Science Classroom." Paper presented at the Fifth Conference on Computers in the Undergraduate Curricula, Washington State University, Pullman, Washington.

Parker, S. and M. Ross (1977) "Experiences with SETUPS: American Politics -- A Comparative Evaluation of Simulations." DEA News 12. (A publication of the Division of Educational Affairs of the American Political Science Association)

Rywick, T. (1975) "Increasing Student Interest by the Use of Interactive Computer Simulations." Behavior Research Methods and Instrumentation 7.

Stout, R. L. (1974) "Modeling and the Michigan Experimental Simulation Supervisor: An Overview and some Prospects." Behavior Research Methods and Instrumentation 6.

Weiss, C. H. (1972) Evaluation Research: Methods of Assessing Program Effectiveness. Englewood Cliffs, New Jersey: Prentice-Hall.